k-Means (Family)

MP

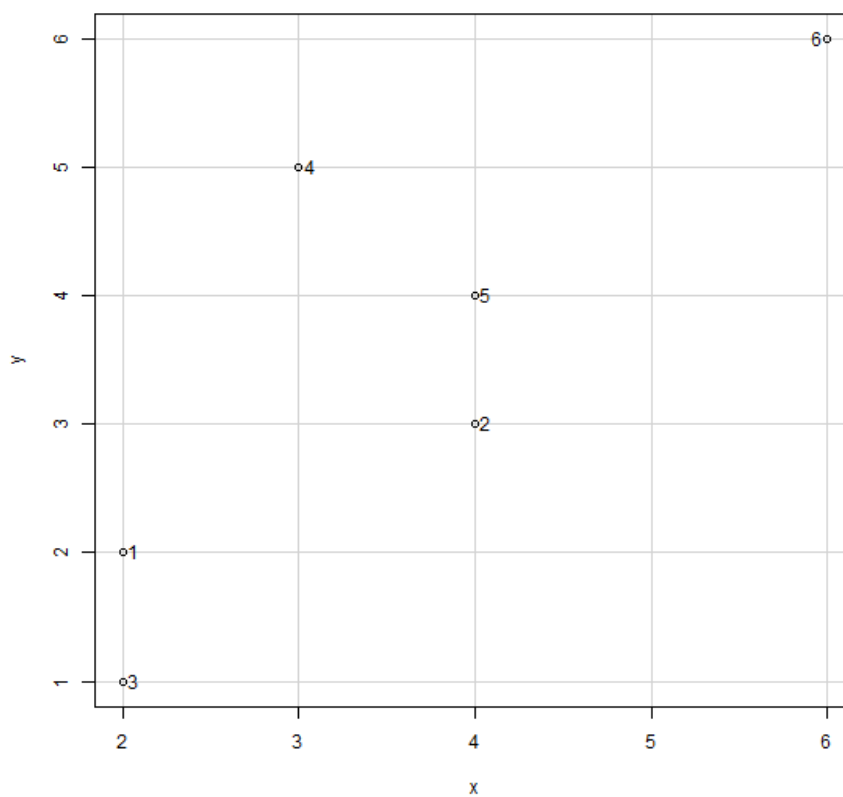### 2015-06-23

```
> setwd("D:/Dropbox/R/2015-NUS/Session-3/(b) k-Means/Family")
```

```
> Dataset <-
+   read.table("D:/Dropbox/R/2015-NUS/Session-3/(b) k-Means/Family/Families-k.csv",
+     header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
```

```
> scatterplot(y~x, reg.line=FALSE, smooth=FALSE, spread=FALSE,
+   id.method='mahal', id.n = 6, boxplots=FALSE, span=0.5, data=Dataset)
```



```
1 2 3 4 5 6
1 2 3 4 5 6
```

```
> .cluster <-  KMeans(model.matrix(~-1 + x + y, Dataset), centers = 3,
+   iter.max = 10, num.seeds = 10)
> .cluster$size # Cluster Sizes
```

```
[1] 3 1 2
```

```
> .cluster$centers # Cluster Centroids
```

```
    new.x.x new.x.y
1 3.666667     4.0
2 6.000000     6.0
3 2.000000     1.5
```

```
> .cluster$withinss # Within Cluster Sum of Squares
```

```
[1] 2.666667 0.000000 0.500000
```

```
> .cluster$tot.withinss # Total Within Sum of Squares
```

```
[1] 3.166667
```

```
> .cluster$betweenss # Between Cluster Sum of Squares
```

```
[1] 25.83333
```

```
> Dataset$KMeans <- assignCluster(model.matrix(~-1 + x + y, Dataset), Dataset,
+     .cluster$cluster)
> remove(.cluster)
```